

SPECIFICATION

Method for Speech Recognition, Apparatus for the Same, and Voice
Controller

FIELD OF THE INVENTION

The present invention relates to a method and an apparatus for recognizing a speaker independent speech, and a voice controller including the speech recognition apparatus.

BACKGROUND OF THE INVENTION

Speech recognition methods are disclosed in Transaction of The Institute of Electronics and Communication Engineers of Japan. Vol. J63-D No. 12 pp. 1002-1009, December, 1980 and Japanese Patent Application Non-examined Publication No. H10-282986. In these speech recognition methods, speakers are previously classified by characteristics such as their ages to trained patterns.

A speaker adaptation method is also widely studied in Wakita, H. "Normalization of Vowels by Vocal-Tract Length and Its Application to Vowel Identification," IEEE (Institute of Electrical and Electronics Engineers) Trans. ASSP 25 (2): pp. 183-192 (1977). This speaker adaptation method distorts a spectral frequency of a speech sound of a speaker by using a single pattern.

A maximum a posteriori estimation (MAP estimation) or the like is known as a speaker adaptation method capable of assimilating a detailed characteristic of a speaker. Technical Report of IEICE (The Institute of Electronics, Information and Communication Engineers) Vol. 93 No. 427 pp. 39-46 (SP93-133, 1993) discloses the MAP estimation.

This method, however, has a problem that if training utterances as a sample beforehand accumulated for an adaptation are extremely few, for example, using only one utterance is spoken, the adaptation cannot improve speech recognition.

A method having a higher recognition rate of a speaker independent word recognizer is disclosed in, for example, Japanese Patent Application Non-examined Publication No. H5-341798. In this speech recognition method, a speaker speaks one of names being given to a speech recognition

apparatus, and the apparatus selects a database adequate to the speaker based on the speech sounds. After that, the speaker speaks a word to be recognized, and the word is processed by speech recognition using the selected database.

- 5 This method, however, has a problem that it is necessary to always examine firstly whether or not the utterance of the speaker is the name of the device, and therefore it takes time for processing. Additionally, this conventional apparatus simply selects databases to be used for a next utterance based on the discrimination whether or not the speaker is adapted, so that a large memory capacity for storing the databases is required.

In the prior art discussed above, detailed characteristics of a speaker are hardly assimilated based on a few utterances, namely only one word or several words at the most, which results in insufficient speech recognition performance.

- 15 It is an object of the present invention to improve speech recognition performance by assimilating detailed characteristics of a speaker based on a few utterances even if a memory capacity for storing databases is small.

SUMMARY OF THE INVENTION

- 20 The present invention addresses the problems discussed above, and aims to provide a speech recognition method which comprises the steps of:

selecting, based on a first utterance by a speaker, an adaptable trained pattern from a plurality of trained patterns that are classified by the characteristics of training speakers who speaks training utterances;

- 25 finding a distortion coefficient fixed by spectral region of speech for a utterance by the speaker based on the selected trained pattern and a first utterance by the speaker; and

recognizing an input utterance following the first utterance using the selected trained pattern and the distortion coefficient.

- 30 A speech recognition apparatus in coincidence with the present invention comprises the following elements:

(a) an acoustic analysis unit for acoustically analyzing an input speech sound to provide acoustic parameters;

- 35 (b) a pattern by-characteristic storage for previously holding a plurality of trained patterns classified by characteristics of training speakers;

(c) a pattern by-characteristic selection unit for selecting an adaptable trained pattern from the plurality of trained patterns based on a

first utterance by a speaker;

(d) a speaker adaptation processor for obtaining a distortion coefficient fixed by spectral region of speech for acoustic parameters of the first utterance using the acoustic parameters and the trained pattern selected by the pattern selection unit;

(e) a word lexicon including known words to be recognized; and

(f) a speech recognition unit for recognizing an input speech sound following the first utterance using the distortion coefficient, the selected trained pattern, and the word lexicon.

A voice controller in coincidence with the present invention comprises the following elements:

(a) a sound input unit for receiving speech sounds;

(b) an acoustic analysis unit for acoustically analyzing a speech sound from the sound input unit to provide acoustic parameters;

(c) a pattern by-characteristic storage for previously holding a plurality of trained patterns classified by characteristics of training speakers;

(d) a pattern by-characteristic selection unit for selecting an adaptable trained pattern from the plurality of trained patterns based on a first utterance by a speaker;

(e) a speaker adaptation processor for determining a distortion coefficient fixed by spectral region of speech for acoustic parameters of the first utterance, using the acoustic parameters and the trained pattern selected by the pattern selection unit;

(f) a word lexicon including known words to be recognized; and

(g) a speech recognition unit for recognizing an input speech sound following the first utterance using the distortion coefficient, the selected trained pattern, and the word lexicon; and

(h) a control signal output unit for outputting a control signal based on a recognition result supplied from the speech recognition unit.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram of a voice control system in coincidence with a first exemplary embodiment of the present invention.

Fig. 2 is a block diagram of a voice controller in coincidence with the first exemplary embodiment.

Fig. 3 is a detailed block diagram of a pattern by-characteristic storage in coincidence with the first exemplary embodiment.